

# QMA variants with polynomially many provers

Sevag Gharibian\*

Jamie Sikora<sup>†</sup>

Sarvagya Upadhyay<sup>‡</sup>

August 3, 2011

## Abstract

We study three variants of multi-prover quantum Merlin-Arthur proof systems. We first show that the class of problems that can be efficiently verified using polynomially many quantum proofs, each of logarithmic-size (denoted  $\text{QMA}_{\log}(\text{poly})$ ), is exactly MQA, the class of problems which can be efficiently verified via a classical proof and a quantum verifier. Assuming  $\text{BQP} \neq \text{MQA}$ , this achieves the separation  $\text{QMA}_{\log}(1) \neq \text{QMA}_{\log}(\text{poly})$ . We then study the class  $\text{BellQMA}(\text{poly})$ , characterized by a verifier who first applies unentangled, nonadaptive measurements to each of the polynomially many proofs, followed by an arbitrary but efficient quantum verification circuit on the resulting measurement outcomes. We show that if the number of outcomes per nonadaptive measurement is a polynomially-bounded function, then the expressive power of the proof system is exactly QMA. Finally, we study a class equivalent to  $\text{QMA}(m)$ , denoted  $\text{SepQMA}(m)$ , where the verifier's measurement operator corresponding to outcome *accept* is a fully separable operator across the  $m$  quantum proofs. Using cone programming duality, we give an alternate proof of a result of Harrow and Montanaro [FOCS, p. 633–642 (2010)] that shows a perfect parallel repetition theorem for  $\text{SepQMA}(m)$  for any  $m$ .

## 1 Introduction and summary of results

The study of classical proof systems has yielded some of the greatest achievements in theoretical computer science, from the Cook-Levin theorem [Coo71, Lev73], which formally ushered in the age of NP verification systems and the now ubiquitous notion of NP-hardness, to the more modern PCP theorem [AS98, ALM<sup>+</sup>98], which has led to significant advancements in our understanding of hardness of approximation. A natural generalization of the class NP, or more accurately its probabilistic cousin Merlin-Arthur (MA), to the quantum setting is the class quantum Merlin-Arthur (QMA) [KSV02], where a computationally powerful but untrustworthy prover, Merlin, sends a *quantum* proof to convince an efficient *quantum* verifier, Arthur, that a given input string  $x \in \{0, 1\}^n$  is a YES-instance for a specified promise problem.

More specifically, a QMA proof system for a given promise problem  $A$  is characterized by the following properties (see Section 2.1 for formal definitions):

- For every YES-instance  $x$  of  $A$ , there exists a polynomial-size quantum proof which can convince Arthur of this fact with high probability, with the smallest such success probability over all YES-instances called the *completeness* of the protocol.

---

\*David R. Cheriton School of Computer Science and Institute for Quantum Computing, University of Waterloo. Address: 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada. Email: sggharib@cs.uwaterloo.ca.

<sup>†</sup>Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo. Address: 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada. Email: jwsikor@uwaterloo.ca.

<sup>‡</sup>David R. Cheriton School of Computer Science and Institute for Quantum Computing, University of Waterloo. Address: 200 University Ave. W., Waterloo, ON, N2L 3G1, Canada. Email: supadhya@cs.uwaterloo.ca .

- For every NO-instance  $x$  of  $A$  and for *any* purported quantum proof, Arthur rejects with high probability, with the maximum success probability over all NO-instances called the *soundness* of the protocol.

It is easy to see that QMA proof systems are at least as powerful as NP or MA, since the ability to process and exchange quantum information does not prevent Arthur from choosing to act classically.

Much attention has been devoted to QMA over recent years. We now have a number of problems which are *complete* for QMA (see e.g. [Bra06, Liu06, BS07, LCV07, SV09, JGL10, WMN10, Ros11]), with the quantum analogue of classical constraint satisfaction, the physically motivated  $k$ -local Hamiltonian problem [KSV02, KR03, KKR06, OT08, AGIK09], being the canonical QMA-complete problem. In analogy with NP-complete problems, it is tempting to think of QMA-complete problems as hard even for a *quantum* computer to solve, though this is somewhat of a misnomer as even NP-complete problems are generally believed to be intractable for quantum computers. QMA is an extremely robust complexity class that satisfies strong error-reduction properties, and using these properties one can, e.g., give a very elegant and simple proof that  $\text{MA} \subseteq \text{QMA} \subseteq \text{PP}$  (the first containment follows trivially from the definition) [MW05]. However, there still remain important open questions — for example, despite the fact that MA is contained in the polynomial hierarchy (PH) [AB09], we do not know whether  $\text{QMA} \subseteq \text{PH}$ .

An approach for understanding a complexity class is to consider how introducing variations to its definition changes its properties. In this paper, we thus ask: *How does allowing multiple unentangled provers affect the expressive power of QMA?* In particular, we are interested in variants of the class  $\text{QMA}(\text{poly})$ , a.k.a. quantum Merlin-Arthur proof systems with polynomially many Merlins, where the verifier receives a polynomial number of quantum proofs, which are promised to be unentangled with each other. Note that the *classical* version of this class collapses trivially to MA, as the set of potential strategies of a single Merlin and the set of potential strategies of multiple Merlins coincide. This logic fails, however, in the quantum case, as a single Merlin simulating the action of multiple Merlins can try to cheat by entangling the multiple proofs. Despite much effort, very little is known (more details under *Previous Work* below) about the structural properties of  $\text{QMA}(\text{poly})$ , except for the obvious containments  $\text{QMA} \subseteq \text{QMA}(\text{poly}) \subseteq \text{NEXP}$ .

**Our results:** We show the following three results regarding variants of  $\text{QMA}(\text{poly})$ .

**1. Relationship to MQA.** We first show that restricting the definition of  $\text{QMA}(\text{poly})$  so that each prover’s proof is at most a *logarithmic* number of qubits (a class which we henceforth denote as  $\text{QMA}_{\log}(\text{poly})$ ) collapses it to the class MQA, where MQA is defined<sup>1</sup> as QMA except that Merlin’s proof is a polynomial-size *classical* string. In other words, if each prover is restricted to sending short quantum proofs, then one can not only do away with multiple provers, but also of the need for *quantum* proofs altogether. Specifically, we show:

**Theorem 1.1.**  $\text{QMA}_{\log}(\text{poly}) = \text{MQA}$ .

The significance of this result is as follows: Understanding the expressive power of  $\text{QMA}(\text{poly})$ , or its relationship with QMA, is currently one of the biggest challenges in quantum complexity theory. (It was recently shown that, in fact,  $\text{QMA}(\text{poly}) = \text{QMA}(2)$  [HM10], where  $\text{QMA}(2)$  is QMA with two unentangled provers.) Theorem 1.1 settles this question in the logarithmic-size

---

<sup>1</sup>Note: MQA has been studied [AN02, JW06, Aar06, AK07, Bei08, ABOBS08, WY08] under the name QCMA in the literature — the notation MQA was suggested by Watrous [Wat09].

message setting. Moreover, it immediately implies the following separation, where  $\text{QMA}_{\log}$  is  $\text{QMA}$  restricted to a logarithmic-size proof:

**Corollary 1.2.** *If  $\text{BQP} \neq \text{MQA}$ , then  $\text{QMA}_{\log} \neq \text{QMA}_{\log}(\text{poly})$ .*

This follows simply because  $\text{QMA}_{\log} = \text{BQP}$  [MW05]. Note that the assumption  $\text{BQP} \neq \text{MQA}$  is reasonable, as otherwise quantum computers could efficiently solve NP-complete problems.

**2. Towards a non-trivial upper bound on  $\text{BellQMA}(\text{poly})$ .** Another approach to studying the question of whether  $\text{QMA} = \text{QMA}(\text{poly})$  is to understand the properties of restricted versions of  $\text{QMA}(\text{poly})$ , and this is precisely where the class  $\text{BellQMA}(\text{poly})$  comes into play.  $\text{BellQMA}(\text{poly})$  is defined [Bra08, ABD<sup>+</sup>09, CD10] analogously to  $\text{QMA}(\text{poly})$ , except that before applying his quantum verification circuit to the polynomially many unentangled quantum proofs, Arthur must measure each proof using a nonadaptive and unentangled (across all proofs) measurement (we call this *Stage 1* of the verification). He then feeds the resulting *classical* outcomes induced by these measurements into his arbitrary efficient quantum circuit (we call this *Stage 2*). This quantum circuit implements a two-outcome measurement operation corresponding to outcomes *accept* and *reject*.

The significance of  $\text{BellQMA}(\text{poly})$  in our setting is that if one could show that  $\text{QMA} \neq \text{BellQMA}(\text{poly})$ , then it would follow that  $\text{QMA} \neq \text{QMA}(\text{poly})$ , since  $\text{QMA} \subseteq \text{BellQMA}(\text{poly}) \subseteq \text{QMA}(\text{poly})$ . To this end, Brandão has shown the negative result that  $\text{QMA} = \text{BellQMA}(m)$  for *constant*  $m$  [Bra08]. Where  $\text{BellQMA}(\text{poly})$  lies, however, remains open. For example, although we know  $\text{QMA}(2) = \text{QMA}(\text{poly})$  [HM10], the same techniques do not apply in any obvious way to show an analogous result  $\text{BellQMA}(2) = \text{BellQMA}(\text{poly})$  as they require entangled measurements (i.e. SWAP test measurements) across multiple proofs, which violate the definition of  $\text{BellQMA}$ .

To make progress on  $\text{BellQMA}(\text{poly})$ , we introduce the class  $\text{BellQMA}[r, m]$ , which is defined as  $\text{BellQMA}(m)$  with  $m$  provers and the additional restriction that in Stage 1 above, the number of outcomes per proof in Arthur’s nonadaptive measurements is upper bounded by  $r$ . Our contribution is the following:

**Theorem 1.3.** *For any polynomially bounded functions  $r, m : \mathbb{N} \rightarrow \mathbb{N}$ ,  $\text{BellQMA}[r, m] = \text{QMA}$ .*

In other words,  $\text{BellQMA}(\text{poly})$  cannot be used to show  $\text{QMA} \neq \text{QMA}(\text{poly})$  if the verifier in the  $\text{BellQMA}(\text{poly})$  protocol is restricted to have a polynomially bounded number of measurement outcomes per proof in Stage 1. We remark that, in general, the number of such measurement outcomes can be exponential in the input length — the restriction that  $r$  be a polynomially bounded function is crucial for the proof of Theorem 1.3. For this reason, our result complements, rather than subsumes Brandão’s result [Bra08]. In other words, in our notation, Brandão has shown that  $\text{BellQMA}[\text{exp}, \text{const}] = \text{QMA}$ , and we show  $\text{BellQMA}[\text{poly}, \text{poly}] = \text{QMA}$ .

Readers should note that we allow the second stage of the verification procedure to be *quantum*, as per the definition suggested by Chen and Drucker [CD10], as opposed to *classical*, as studied by Brandão [Bra08]. We remark that the conclusion of Theorem 1.3 holds even if the second stage of verification is completely classical.

Finally, it is worth noting that by combining Theorems 1.1 and 1.3, we conclude that in the setting of  $\text{BellQMA}(\text{poly})$ , if  $\text{MQA} \neq \text{QMA}$ , then having the Merlins send logarithmic-size proofs without any restriction on the number of local measurement outcomes of Arthur in Stage 1 has less expressive power than sending polynomial-size proofs but restricting the number of outcomes, even though the number of measurement outcomes in Stage 1 per Merlin in both cases is the same, i.e. polynomial in the input length.

**3. Perfect parallel repetition for SepQMA( $m$ ).** A key question in designing proof systems is how to improve the completeness and soundness parameters of a verification protocol without increasing the required number of rounds of communication. A natural approach for doing so is to repeat the protocol multiple times in parallel. With QMA, however, this raises the concern that Merlin might try to cheat by entangling his proofs across these parallel runs. If, though, *perfect parallel repetition* holds, it means that for any input string  $x$ , if the verification procedure  $V$  accepts with probability  $p(|x|)$ , then if we run  $V$   $k$  times in parallel, the probability of accepting in all  $k$  runs of  $V$  is precisely  $p(|x|)^k$ . Note that we do not put any restriction on the quantum proof, which can be entangled across the  $k$  executions of the protocol. In other words, if perfect parallel repetition holds, there is no incentive for Merlin to cheat — an honest proof which is a product state across all  $k$  runs achieves the maximum success probability.

Our final contribution is an alternate proof of a perfect parallel repetition theorem for a class which is equivalent [HM10] to QMA( $m$ ), namely SepQMA( $m$ ). The theorem was first proved in Harrow and Montanaro [HM10] in connection with an error reduction technique for QMA(poly). However, our proof is significantly different from theirs and uses the cone programming characterization of QMA(poly). Here SepQMA( $m$ ) is defined as QMA( $m$ ) with the restriction that Arthur’s measurement operator corresponding to acceptance is a *separable* operator across the  $m$  unentangled proofs. (Note that this does not imply that Arthur’s measurement operator corresponding to rejection is also separable.) We show:

**Theorem 1.4** (see [HM10] for alternate proof). *The class SepQMA( $m$ ) admits perfect parallel repetition.*

Our alternate proof of Theorem 1.4 is significant in that, to the best of our knowledge, it is the first use of duality theory for a cone program *other* than a semidefinite program to establish a parallel repetition result (note that cone programming generalizes semidefinite programming). We remark that semidefinite programs have been previously used to show perfect or strong parallel repetition theorems for various other models of (single or two-prover) quantum interactive proof systems [CSUU08, KRT10, Gut09], and that the alternate proof of Theorem 1.4 of Harrow and Montanaro is not based on semidefinite programming. Perfect parallel repetition for SepQMA( $m$ ) in itself is interesting, as it has been used to show that error reduction is possible for QMA( $m$ ) proof systems [HM10].

**Proof ideas and tools:** The proof of our first result, Theorem 1.1, is a simple application of the fact that quantum states of logarithmic-size can be described to within inverse exponential precision using a polynomial number of bits and efficiently prepared by a quantum circuit. Hence, roughly speaking, one can replace a polynomial number of logarithmic-size quantum proofs with a single polynomial size classical proof. Each quantum proof can then be efficiently prepared from this classical information by the verifier to within inverse exponential precision since the original quantum proofs were of logarithmic-size. Although the proof is simple, one cannot hope for a better characterization using other techniques because the reverse containment, i.e.  $\text{MQA} \subseteq \text{QMA}_{\log}(\text{poly})$ , also holds using similar ideas.

More technically challenging is our second result, Theorem 1.3. To show the containment  $\text{BellQMA}[\text{poly}, \text{poly}] \subseteq \text{QMA}$  (note that the reverse containment  $\text{QMA} \subseteq \text{BellQMA}[\text{poly}, \text{poly}]$  is trivial since  $\text{QMA} \subseteq \text{BellQMA}[2, 1]$ ), we demonstrate a QMA protocol which simulates an arbitrary BellQMA[poly, poly] protocol using the following observation: although consolidating  $m$  quantum proofs into a single quantum proof raises the possibility of cheating using entanglement, if Arthur is also sent an appropriate classical “consistency-check” string, then a dishonest Merlin can be caught with non-negligible probability.



Specifically, in our QMA protocol, we ask a single Merlin to send the  $m$  quantum proofs of the original BellQMA protocol (denoted by a single state  $|\psi\rangle$ ), accompanied by a “consistency-check” string  $\mathbf{p}$  which is a classical description of the probability distributions obtained as the output of Stage 1. One can think of this as having the QMA verifier *delegate* Stage 1 of the BellQMA verification to Merlin. Arthur then performs a consistency check between  $|\psi\rangle$  and  $\mathbf{p}$  based on the premise that if Merlin is honest, then  $\mathbf{p}$  should arise from running Stage 1 of the original verification on  $|\psi\rangle$ . If this check passes, then Arthur runs Stage 2 of the BellQMA verification on  $\mathbf{p}$ . If Merlin tries to cheat, however, we show that the check detects this with non-negligible probability. Note that the accuracy of the consistency check crucially uses the fact that there are at most polynomially many outcomes to check for each local measurement of Stage 1.

Finally, our last result, Theorem 1.4, is shown using duality theory for a class of cone programs that captures the success probability of a QMA(poly) protocol. In particular, we phrase the maximum acceptance probability of a (possibly cheating) prover for the two-fold repetition of a SepQMA( $m$ ) verification protocol as a cone program. We then demonstrate a feasible solution for its dual yielding an upper bound on the maximum acceptance probability. The objective value of this dual solution is precisely the product of the optimum values of the two instances of the SepQMA( $m$ ) verification protocols. We conclude that one of the optimal strategies of the provers is to be faithful in the following sense: Each prover elects not to entangle his/her two quantum proofs for the two instances of the SepQMA( $m$ ) protocol and instead sends a tensor product of optimal proofs for both the instances.

**Previous work.** The expressive power of multiple Merlins was first studied by Kobayashi, Matsumoto and Yamakami [KMY03], who showed that  $\text{QMA}(2) = \text{QMA}(\text{poly})$  if and only if the class of QMA(2) protocols with completeness  $c$  and soundness  $s$  (with at least inverse polynomial gap) is exactly equal to QMA(2) protocols with completeness  $2/3$  and soundness  $1/3$ . A substantial amount of research has since been devoted to understanding the properties of multi-prover quantum Merlin-Arthur proof systems. Recently, Harrow and Montanaro [HM10] demonstrated a *product state test*, wherein given two copies of a *pure* quantum state on multiple systems, the test distinguishes between the cases when the quantum state is a *fully* product state across all the systems or *far* from any such state. Using this test, they answered a few important questions regarding QMA(poly). In particular, they showed that

$$\text{QMA}(2) = \text{QMA}(\text{poly})$$

and that error reduction is possible for such proof systems. Prior to their result, the answers to both the questions were known to be affirmative assuming a *weak* version of the Additivity Conjecture [ABD<sup>+</sup>09]. One of the crucial properties of the product state test is that it can be converted into a QMA(2) protocol, where Arthur’s measurement operator corresponding to outcome *accept* is a separable operator across the two proofs. Harrow and Montanaro established a perfect parallel repetition theorem for such proof systems, a crucial step in obtaining exponentially small error probabilities.

Blier and Tapp initiated the study of *logarithmic*-size unentangled quantum proofs [BT09]. They showed that two unentangled quantum proofs suffice to show that a 3-coloring of an input graph exists, implying that NP has *succinct* unentangled quantum proofs. A drawback of their protocol is that although it has *perfect* completeness, its soundness is only inverse polynomially bounded away from 1. Shortly after, Aaronson, Beigi, Drucker, Fefferman and Shor [ABD<sup>+</sup>09] showed that satisfiability of any 3-SAT formula of size  $n$  can be proven by  $\tilde{O}(\sqrt{n})$  unentangled quantum proofs of  $O(\log n)$  qubits with perfect completeness and constant soundness (see also [CD10]). In

a subsequent paper [Bei08], Beigi improved directly on Blier and Tapp's result [BT09] by showing that by sacrificing perfect completeness, one can show that NP has two logarithmic-size quantum proofs with a better gap between completeness and soundness probabilities than in [BT09].

Finally, one of the open questions raised in Ref. [ABD<sup>+</sup>09] concerns the power of Arthur's verification procedure. In particular, the paper introduces two different classes of verification procedures, BellQMA and LOCCQMA verification. Roughly speaking, LOCCQMA verification corresponds to Arthur applying a measurement operation that can be implemented by Local Operations and Classical Communication (LOCC) (with respect to the partition induced by the multiple proofs). The authors raised the question of whether  $\text{BellQMA}(\text{poly}) = \text{QMA}$  or not. Brandão [Bra08] showed that  $\text{BellQMA}(m)$  is equal to QMA for constant  $m$ . In a recent development, Brandão, Christandl and Yard [BCY11] showed that  $\text{LOCCQMA}(m)$  is equal to QMA for constant  $m$ .

**Organization of this paper.** We begin in Section 2 with background and notation, defining relevant complexity classes in Section 2.1, and reviewing cone programming in Section 2.2. Theorems 1.1, 1.3, and 1.4 are proved in Sections 3, 4, and 5, respectively. We conclude with open problems in Section 6.

## 2 Preliminaries and Notation

We begin by setting our notation, and subsequently review the background material required for this paper. First, the notation  $[m]$  indicates the set  $\{1, \dots, m\}$ , and  $|x|$  the length of a string  $x \in \{0, 1\}^*$ . We let uppercase script letters  $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$  denote complex Euclidean spaces. We denote the sets of linear, Hermitian, positive semidefinite, and density operators acting on vector space  $\mathcal{X}$  by  $\text{L}(\mathcal{X})$ ,  $\text{Herm}(\mathcal{X})$ ,  $\text{Pos}(\mathcal{X})$ , and  $\text{D}(\mathcal{X})$ , respectively. We denote the standard Hilbert-Schmidt inner product of operators  $A$  and  $B$  as  $\langle A, B \rangle := \text{Tr}(A^*B)$ , where  $A^*$  denotes the adjoint of  $A$ . The spectral and trace norms of an operator  $A$  are given by  $\|A\|_\infty := \max\{\|Au\| : \|u\| = 1\}$  and  $\|A\|_{\text{tr}} := \text{Tr}(\sqrt{A^*A})$ , respectively, where  $\|u\|$  denotes the Euclidean norm of a vector  $u$ . These can be thought of as the largest singular value and the sum of singular values of  $A$ , respectively. A useful lemma in this paper regarding the trace norm is the following:

**Lemma 2.1** ([Wat02]). *Let  $\{\rho_1, \dots, \rho_k\} \subset \text{D}(\mathcal{X})$  and  $\{\sigma_1, \dots, \sigma_k\} \subset \text{D}(\mathcal{X})$ . Then*

$$\left\| \bigotimes_{i=1}^k \rho_i - \bigotimes_{i=1}^k \sigma_i \right\|_{\text{tr}} \leq \sum_{i=1}^k \|\rho_i - \sigma_i\|_{\text{tr}}.$$

Next, we say a (possibly unnormalized) operator  $A \in \text{Pos}(\mathcal{X}_1 \otimes \dots \otimes \mathcal{X}_m)$  is *fully separable* if it can be written as

$$A = \sum_{i=1}^k P_1(i) \otimes \dots \otimes P_m(i),$$

where  $P_j(i) \in \text{Pos}(\mathcal{X}_j)$ , for every  $j \in [m]$  and  $i \in [k]$ . The set of fully separable operators is denoted  $\text{Sep}(\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m)$ . This notation is helpful in the context of cone programming. In the setting of quantum information, one typically also has  $\text{Tr}(A) = 1$ . The set of fully separable density operators is convex, compact, and has non-empty interior since it contains a ball around the normalized identity operator [GB02, GB03, GB05]. Using these facts, one can formally prove that the problem of determining whether a given density operator is *close* to the set of separable states is NP-hard for the bipartite case (i.e.,  $m = 2$ ) [Gur03, Ioa07, Gha10, Bei08]. For the multipartite case

(i.e.,  $m > 2$ ), the classification of entanglement becomes much more intricate due to, e.g., notions of *partial separability* (hence our use of the term *fully separable*) — see the reference [HHHH09] for a comprehensive survey.

We use the fact that any pure quantum state  $|\psi\rangle \in \mathbb{C}^N$  can be described approximately classically using  $N \cdot f(N)$  bits, for some function  $f : \mathbb{N} \rightarrow \mathbb{N}$ . The resulting approximate description  $|\psi'\rangle$  satisfies  $\| |\psi\rangle - |\psi'\rangle \| \leq N2^{-(f(N)+1)}$ . We also speak in terms of *quantum registers* rather than quantum states in the next two sections. To make the association precise, an  $n$ -qubit quantum register  $X$  is associated with a vector space  $\mathcal{X} = \mathbb{C}^{2^n}$  and contains any element of  $D(\mathcal{X})$ .

Finally, moving to quantum operations, the notion of measurement used in this paper is that of a Positive Operator Valued Measure (POVM), given by a finite set of positive semidefinite operators  $\{\Pi_1, \dots, \Pi_r\} \subset \text{Pos}(\mathcal{X})$  obeying

$$\sum_{i=1}^r \Pi_i = \mathbb{1}_{\mathcal{X}}.$$

Regarding unitary operators, we use the fact that any unitary operator acting on  $k$  qubits can be approximated within high precision by a finite set of one-qubit, two-qubit, and/or three-qubit unitary operators. Such a finite set is often referred to as an *approximately* universal set of quantum gates, and one such set is comprised of the Toffoli, Hadamard, and phase-shift gates. The Solovay-Kitaev theorem implies that the action of an arbitrary unitary operator  $U$  on  $k$  qubits can be simulated by a composition  $\tilde{U}$  of  $O(4^k \log(1/\epsilon))$  universal gates, such that  $\|U - \tilde{U}\|_{\infty} \leq \epsilon$  [NC00].

## 2.1 Relevant quantum complexity classes

A promise problem  $A = (A_{\text{yes}}, A_{\text{no}})$  is a partition of the set  $\{0,1\}^*$  into three disjoint subsets: the set  $A_{\text{yes}}$  denotes the set of YES-instances of the problem, the set  $A_{\text{no}}$  denotes the set of NO-instances of the problem, and the set  $\{0,1\}^* \setminus (A_{\text{yes}} \cup A_{\text{no}})$  is the set of disallowed strings (we are *promised* the input does not fall into this last set).

**Definition 2.2** (QMA( $m$ )). *Let  $p : \mathbb{N} \rightarrow \mathbb{N}$  be a polynomially bounded function, and  $m : \mathbb{N} \rightarrow \mathbb{N}$  a function. A promise problem  $A = (A_{\text{yes}}, A_{\text{no}})$  is in the class QMA( $m$ ) if there exists a polynomial-time generated family of verification circuits  $Q = \{Q_n \mid n \in \mathbb{N}\}$  with the following properties:*

1. *Each  $Q_n$  acts on  $n + p(n)$  input qubits, and outputs one qubit.*
2. *(Completeness) For every  $x \in A_{\text{yes}}$ , there exist  $m(|x|)$  quantum proofs  $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_{m(|x|)}\rangle \in \mathbb{C}^{2^{p(|x|)}}$  such that*

$$\Pr[Q_{|x|} \text{ accepts } (x, |\psi_1\rangle \otimes \dots \otimes |\psi_{m(|x|)}\rangle)] \geq 2/3.$$

3. *(Soundness) For any  $x \in A_{\text{no}}$  and any  $m(|x|)$  quantum proofs  $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_{m(|x|)}\rangle \in \mathbb{C}^{2^{p(|x|)}}$*

$$\Pr[Q_{|x|} \text{ accepts } (x, |\psi_1\rangle \otimes \dots \otimes |\psi_{m(|x|)}\rangle)] \leq 1/3.$$

Furthermore, the class QMA(poly) is defined as  $\text{QMA}(\text{poly}) = \bigcup_{m \in \text{poly}} \text{QMA}(m)$ .

We remark that the constants  $2/3$  and  $1/3$  can be replaced by any  $a, b \geq 0$ , respectively, such that  $a - b \geq 1/\text{poly}(n)$ . This does not change the expressive power of the proof system. All complexity classes considered in this paper are variants of QMA( $m$ ) and satisfy the properties mentioned above in Definition 2.2. We define the following classes, which are relevant to this paper.

1. **[QMA and MQA]** The class QMA is simply QMA(1). If we replace the quantum proofs in the definition of QMA with a polynomial-size classical proof string, the corresponding class is denoted MQA.
2. **[SepQMA(poly)]** The class SepQMA(poly) is a subclass of QMA(poly), wherein Arthur's measurement operator corresponding to outcome *accept* is a fully separable operator across the proofs.
3. **[QMA<sub>log</sub>(poly)]** The class QMA<sub>log</sub>(poly) is a subclass of QMA(poly), wherein each Merlin's message to Arthur is  $O(\log(|x|))$  qubits in length.

Finally, for clarity, we opt to give a more formal definition of BellQMA[ $r, m$ ].

**Definition 2.3** (BellQMA[ $r, m$ ]). *Let  $r, m : \mathbb{N} \rightarrow \mathbb{N}$  be two functions. A promise problem  $A = (A_{\text{yes}}, A_{\text{no}})$  is in class BellQMA[ $r, m$ ] if there exists a QMA( $m$ ) verification protocol in which Arthur is restricted to act as follows.*

1. *Arthur performs a polynomial-time quantum computation on the input  $x$  and generates a description of quantum circuits  $V_1(x), \dots, V_m(x)$ , one for each of the  $m$  provers.*
2. *(Stage 1) Arthur simultaneously measures all  $m$  quantum proofs by applying  $V_i(x)$  to the  $i$ -th quantum proof, where the action of  $V_i(x)$  can be described by a unitary operator followed by measurement in the standard basis. The label of the  $i$ -th measurement outcome is stored as a classical string  $y_i$  also identified as an element of  $[r(|x|)]$ .*
3. *(Stage 2) Arthur runs an efficient quantum verification circuit on input  $x$  and measurement outcomes  $(y_1, \dots, y_m)$  to decide whether to accept or reject.*

Note that the key distinction between BellQMA[ $r, m$ ] and BellQMA(poly) is that the former has the number of measurement outcomes in Stage 1 of the protocol bounded by  $r(|x|)$ , whereas the latter may allow exponentially many possible outcomes. Throughout this paper, we use the notation BellQMA[poly, poly] to denote

$$\text{BellQMA}[\text{poly}, \text{poly}] := \bigcup_{r \in \text{poly}} \bigcup_{m \in \text{poly}} \text{BellQMA}[r, m].$$

We remark that, as in [CD10], our BellQMA protocols are allowed to use a *quantum* verification circuit in Stage 2, whereas originally in references [Bra08, ABD<sup>+</sup>09] only classical processing of measurement outcomes  $\{y_i\}$  was allowed in order to emulate the notion of a *Bell experiment* performed by Arthur. We again remark that Theorem 1.3 holds even if Arthur is restricted to do classical processing on the measurement outcomes.

## 2.2 Cone programming

We now briefly review basic notions in conic optimization (or cone programming), which is a generalization of semidefinite optimization. We say that a set  $K$  in an underlying Euclidean space is a cone if  $x \in K$  implies that  $\lambda x \in K$  for all  $\lambda > 0$ . A cone  $K$  is convex if  $x, y \in K$  implies that  $x + y \in K$ . Cone programs are concerned with optimizing a linear function over the intersection of a convex cone and an affine space. It generalizes several well-studied models of optimization including semidefinite programming ( $K = \text{Pos}(\mathcal{X})$ ) and linear programming ( $K = \mathbb{R}_+^n$ ). In this paper, we are primarily concerned with the cone of fully separable operators Sep( $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m$ ) which recall is a closed, convex cone with non-empty interior.



Associated with a cone  $K$  is its dual cone  $K^*$  defined as

$$K^* = \{S : \langle X, S \rangle \geq 0 \text{ for all } X \in K\}.$$

A cone program associates the following 4-tuple  $(C, b, \mathcal{A}, K)$  to an optimization problem described as:

$$\begin{aligned} \text{supremum: } & \langle X, C \rangle \\ \text{subject to: } & \mathcal{A}(X) = b, \\ & X \in K, \end{aligned}$$

where  $\mathcal{A} : \text{Span}(K) \rightarrow \mathbb{R}^m$  is a linear transformation. Note that the inner product is defined as in the Euclidean space. For instance, if the cone under consideration is the set of positive semidefinite or separable operators, then the inner product is the standard Hilbert-Schmidt inner product over the space of Hermitian operators. We say that the cone program is *feasible* if  $\{X : \mathcal{A}(X) = b\} \cap K$  is non-empty and *strictly feasible* if  $\{X : \mathcal{A}(X) = b\} \cap \text{int}(K)$  is non-empty, where  $\text{int}(\cdot)$  denotes the interior of a set.

Cone programs come in primal-dual pairs:

Primal problem (P)	Dual problem (D)
supremum: $\langle X, C \rangle$	infimum: $\langle b, y \rangle$
subject to: $\mathcal{A}(X) = b,$	subject to: $\mathcal{A}^*(y) = C + S,$
$X \in K,$	$S \in K^*,$

where  $\mathcal{A}^*$  is the adjoint of  $\mathcal{A}$ . A convex cone  $K$  is closed if and only if  $K = K^{**}$ . In other words, the dual of the cone  $K^*$  is the original cone  $K$ . Thus, if  $K$  is not closed we need to “order” the primal-dual pairs since  $K \neq K^{**}$  implying the dual of the dual problem is not equal to the primal problem. Since the convex cone of fully separable operators is closed, ordering the primal-dual pairs is not an issue in our case.

Similar to linear programming and semidefinite programming, cone programming has a rich duality theory.

**Lemma 2.4** (Weak Duality). *If  $X$  is primal feasible and  $(y, S)$  is dual feasible then*

$$\langle b, y \rangle - \langle X, C \rangle = \langle X, S \rangle \geq 0.$$

This result can be used to show upper bounds on the value of the primal problem or lower bounds on the value of the dual problem. There is also a notion of *strong duality*. We say that *strong duality holds for a problem (P)* if the optimal value of (P) equals the optimal value of (D) and (D) attains an optimal solution. Below we give a condition that guarantees strong duality for (P).

**Theorem 2.5** (Strong Duality, Version 1). *If (P) is strictly feasible and the optimal value is bounded from above, then strong duality holds for (P), i.e., (D) attains an optimal solution and the optimal values for (P) and (D) coincide.*

In this paper, we are concerned with closed, convex cones with non-empty interior. Since the dual of the dual problem is the primal problem when  $K$  is closed, we can use the following stronger version of strong duality.

**Theorem 2.6** (Strong Duality, Version 2). *Suppose  $K$  is a closed, convex cone. If (P) and (D) are both strictly feasible then strong duality holds for both problems, i.e., both problems attain an optimal solution and the optimal values coincide.*

We refer the reader to the work of Tunçel and Wolkowicz [TW08] and the references therein for more details on cone programming duality.

### 3 Equivalence of MQA and $\text{QMA}_{\log}(\text{poly})$

We now prove Theorem 1.1, i.e., that  $\text{MQA} = \text{QMA}_{\log}(\text{poly})$ . We first show the direction  $\text{MQA} \subseteq \text{QMA}_{\log}(\text{poly})$ . Let  $A = (A_{\text{yes}}, A_{\text{no}})$  be a promise problem in MQA and let  $x \in \{0, 1\}^n$  be the input string. Suppose the MQA prover sends an  $m$ -bit classical proof to the verifier, for polynomially bounded  $m$ . Then the following simple  $\text{QMA}_{\log}(m)$  protocol achieves the desired containment:

#### $\text{QMA}_{\log}(m)$ Protocol

1. **Embed classical bits into qubits.** Each (unentangled) prover  $i \in [m]$  sends a single qubit  $|\psi_i\rangle \in \mathbb{C}^2$  to Arthur. If the  $i$ -th prover is honest, his/her qubit is the computational basis state corresponding to the  $i$ -th bit of the classical MQA proof.
2. **Make things classical again.** Arthur measures all proofs in the computational basis, obtaining a classical string  $y \in \{0, 1\}^m$ .
3. **Run MQA verification.** Arthur runs the MQA verification circuit on  $x$  and  $y$  and accepts if and only if acceptance occurs in the MQA verification.

The completeness property follows straightforwardly. The soundness property is also easy to observe. Note that Arthur runs the MQA verification on a classical string  $y$  and hence he accepts the string with probability at most  $1/3$ .

To show the reverse containment, let  $A = (A_{\text{yes}}, A_{\text{no}})$  be a promise problem in  $\text{QMA}_{\log}(\text{poly})$  and let  $x \in \{0, 1\}^n$  be the input string. Suppose we have a  $\text{QMA}_{\log}(m)$  protocol for polynomially bounded  $m$ , where prover  $i$  sends a  $\lceil c \log n \rceil$ -qubit state  $|\psi_i\rangle$  for some constant  $c > 0$ . Let

$$r(n) = 2^{\lceil c \log n \rceil} = O(n^c).$$

The MQA protocol proceeds as follows:

#### MQA Protocol

1. **Describe proofs classically.** The prover sends  $m$  classical registers represented by the tuple  $(C_1, C_2, \dots, C_m)$ , each of length  $2n \cdot r(n)$  to Arthur. If the prover is honest, register  $C_i$  contains a classical description of the  $i$ -th quantum proof of the  $\text{QMA}_{\log}(m)$  protocol.
2. **State preparation.** Using the contents of register  $C_i$ , for every choice of  $i \in [m]$ , Arthur prepares the state  $|\psi_i\rangle$  by first determining a unitary  $U_i$  such that  $U_i |0 \dots 0\rangle = |\psi_i\rangle$ , and then implementing  $U_i$  with high precision using a finite set of approximately universal gates, obtaining states  $|\psi'_i\rangle$ .
3. **Run  $\text{QMA}_{\log}(m)$  verification.** Arthur runs the  $\text{QMA}_{\log}(m)$  verification circuit on  $|\psi'_1\rangle \otimes \dots \otimes |\psi'_m\rangle$  and accepts if and only if acceptance occurs in  $\text{QMA}_{\log}(m)$  verification.

Observe that each classical register  $C_i$  is of size polynomial in  $n$ , implying the overall proof length is of polynomial size. In Step 1, the prover uses  $n$  bits to represent the real and imaginary parts of each of the polynomially many entities ( $r(n)$  entries) required to describe each  $|\psi\rangle$ . Let the unit vector described by register  $C_i$  be denoted  $|\psi_i\rangle$ . In Step 2,  $U_i$  is easily found as the unitary that maps  $|0 \dots 0\rangle$  to  $|\psi_i\rangle$  as the inverse of the unitary that maps  $|\psi_i\rangle$  to  $|0 \dots 0\rangle$ . Such a unitary can be easily decomposed into a product of polynomially many  $2 \times 2$  rotations on an  $r(n)$ -dimensional real space and a diagonal unitary as follows. The first step is to convert the vector  $|\psi_i\rangle$  into a real vector by applying an appropriate diagonal unitary operator. The second step is to convert

the resulting real unit vector into  $|0 \dots 0\rangle$  by shifting the amplitudes of any standard basis other than  $|0 \dots 0\rangle$  to  $|0 \dots 0\rangle$ . Each of these unitary operators can be implemented by a finite set of approximately universal gates (see Bernstein and Vazirani [BV97] for details). This step also incurs some error, which can be made exponentially small.

Since Steps 1 and 2 can be performed to within inverse exponential error, we thus can ensure  $\|\psi_i\rangle - \psi'_i\rangle\| \leq \epsilon$  for all  $i \in [m]$  and for inverse exponential  $\epsilon > 0$ . By Lemma 2.1, it follows that the overall precision error is at most  $m\epsilon$  for polynomial  $m$ , and thus the completeness and soundness of the protocol are bounded from below and above by (respectively)

$$\frac{2}{3} - m\epsilon \quad \text{and} \quad \frac{1}{3} + m\epsilon.$$

Alternatively, the containment  $\text{QMA}_{\log}(\text{poly}) \subseteq \text{MQA}$  can be shown using a slightly different protocol<sup>2</sup>, where Merlin sends classical descriptions of the quantum circuits that generate the quantum proofs from  $|0 \dots 0\rangle$  instead of classical descriptions of the proofs.

## 4 Equivalence of BellQMA[poly, poly] and QMA

We now show Theorem 1.3, i.e., that  $\text{BellQMA}[r, m] = \text{QMA}$  for polynomially-bounded functions  $r$  and  $m$ . For notational convenience, let  $\Pi_j(i)$  denote Arthur's  $i$ -th POVM element in Stage 1 of the BellQMA verification protocol for the  $j$ -th prover (i.e.  $\sum_{i=1}^r \Pi_j(i) = \mathbb{1}$ ), where we assume without loss of generality that the number of possible outcomes is exactly  $r$  for each prover, and where  $j \in [m]$  for  $m$  the number of provers.

We proceed as follows. Let  $A = (A_{\text{yes}}, A_{\text{no}})$  be a promise problem, and  $x$  be an input string of length  $n := |x|$ . As mentioned in Section 1, the containment  $\text{QMA} \subseteq \text{BellQMA}[\text{poly}, \text{poly}]$  follows straightforwardly since  $\text{QMA} \subseteq \text{BellQMA}[2, 1]$ . For the reverse containment, suppose we have a  $\text{BellQMA}[r, m]$  protocol for polynomially bounded functions  $r, m : \mathbb{N} \rightarrow \mathbb{N}$  with completeness  $2/3$  and soundness  $1/3$ . We show that this protocol can be simulated by a QMA protocol where Merlin sends the following proof to Arthur.

Merlin sends two registers  $(X, Y)$ , which should be thought of as the *classical* and *quantum* registers, respectively. Suppose optimal proofs for the  $\text{BellQMA}[r, m]$  protocol for input  $x$  are given by  $\rho_j$  for  $j \in [m]$ . Then, in the quantum register  $Y$ , an honest Merlin should send many copies of the state  $\rho_j$ . Specifically,  $Y$  is partitioned into  $m$  registers  $Y_j$ , one for each original prover, and each  $Y_j$  should contain  $k$  copies of  $\rho_j$ , for  $k$  a carefully chosen polynomial. In other words,  $Y$  should contain the state  $[\rho_1^{\otimes k}]_{Y_1} \otimes \dots \otimes [\rho_m^{\otimes k}]_{Y_m}$ . We further view each  $Y_j$  as a block of registers  $(Y_j^1, \dots, Y_j^k)$  where  $Y_j^l$  should contain the  $l$ -th copy of  $\rho_j$ .

In the classical register  $X$ , an honest Merlin prepares a quantum state in the computational basis, which intuitively corresponds to a bit string describing the  $m$  classical probability distributions Arthur induces upon applying the measurement operation corresponding to Stage 1 of the BellQMA verification to each of the optimal proofs  $\rho_j$ , respectively.

More formally, we partition  $X$  into  $mr$  registers  $X_j^i$  corresponding to each of the  $j \in [m]$  provers and  $i \in [r]$  POVM outcomes per prover. The content of  $X_j^i$  should be  $p_j(i) := \langle \Pi_j(i), \rho_j \rangle$ , truncated to  $\alpha$  bits of precision ( $\alpha$  polynomially bounded), such that  $\sum_{i=1}^r p_j(i) = 1$ . For example, if the  $j$ -th prover's proof was the single qubit state  $\rho_j = |0\rangle\langle 0|$ , with  $\Pi_j(1) = |0\rangle\langle 0|$  and  $\Pi_j(2) = |1\rangle\langle 1|$ , then  $X_j = (1, 0)$ . We remark that  $X$  plays the role of the classical "consistency check" string described in Section 1.

<sup>2</sup>This protocol was mentioned to us by Richard Cleve.

Of course, Merlin may elect to be dishonest and choose not to send a proof of the above form to Arthur by, e.g., sending a quantum state which is entangled across the registers  $(X, Y)$ . To catch this, our QMA protocol is defined as follows:

### QMA Protocol

1. Merlin sends Arthur a quantum state in registers  $(X, Y)$ , for  $X$  and  $Y$  defined as above.
2. **Force  $X$  to be classical.** Arthur measures register  $X$  in the computational basis and reads the measurement outcome. This forces  $X$  to essentially be a classical register of bits, and destroys any entanglement or correlations between  $X$  and  $Y$ .
3.  **$X$  should contain probability distributions.** Arthur checks whether the content of registers  $X_j$  form a probability distribution  $p_j$ , i.e., that  $\sum_{i=1}^r p_j(i) = 1$ . Arthur rejects if this is not the case.
4. **Consistency check: Can the quantum states in  $Y$  reproduce the distributions in  $X$ ?** Arthur picks independently and uniformly at random, an index  $j \in [m]$  and another index  $i \in [r]$ . He applies the measurement  $\{\Pi_j(i)\}_{i=1}^r$  separately to each register  $Y_j^1, \dots, Y_j^k$ , and counts the number of times outcome  $i$  appears, which we denote henceforth as  $n_j(i)$ . Arthur rejects if

$$\left| \frac{n_j(i)}{k} - p_j(i) \right| \geq \frac{1}{p},$$

for  $p$  a carefully chosen polynomial.

5. **Run Stage 2 of the BellQMA verification and repeat for error amplification.** For each prover  $j$ , Arthur samples an outcome from  $[r]$  according to the distribution in  $(X_j^1, \dots, X_j^r)$ , and runs Stage 2 of the BellQMA verification on the resulting set of samples. He repeats this process independently a polynomial number of times  $q$ , and accepts if and only if the BellQMA procedure accepts on the majority of the runs.

Let us give an intuition behind the above verification procedure. The key step above is Step 4, where Arthur cross-checks that the classical distributions sent in  $X$  really can be obtained by measuring  $m$  quantum proofs, which for an honest Merlin should be unentangled. In this sense, our protocol can alternatively be viewed as using *quantum* proofs ( $Y$ ) to check validity of a *classical* proof ( $X$ ). Intuitively, the reason why entanglement in  $Y$  does not help a dishonest Merlin in Step 3 is due to the local nature of Arthur's checks/measurements. Finally, once Arthur is satisfied that  $X$  contains valid distributions, he runs Step 5. We remark that repetition is used here in order to boost the probability of acceptance in the  $x \in A_{\text{yes}}$  case to exponentially close to 1, which is required to separate it from the  $x \in A_{\text{no}}$  case, where the probability of catching a dishonest Merlin is only inverse polynomially bounded away from 1. Once such a gap exists, standard amplification techniques [KW00, MW05] can be used to further improve completeness and soundness parameters.

To formally analyze completeness and soundness of the protocol, we assign the following values to the parameters mentioned above, all of which are polynomial in  $n$  in our setting:

$$q = 50n \quad \text{and} \quad p = 20mr \quad \text{and} \quad k = 5p^3 \quad \text{and} \quad \alpha = 20nmr.$$

**Completeness.** Intuitively, when  $x \in A_{\text{yes}}$ , Merlin passes Step 4 with probability exponentially close to 1 since he has no incentive to cheat — he can send an unentangled proof in Step 1 to Arthur corresponding to the optimal proofs  $\rho_j$  in the BellQMA protocol, such that the expected value of

$n_j(i)/k$  is indeed  $p_j(i)$ . Arthur's checks in Step 4 are then independent local trials, allowing a Chernoff bound to be applied. We then show that Merlin passes each run in Step 5 with constant probability, and applying the Chernoff bound a second time yields the desired completeness exponentially close to 1 for the protocol.

To state this formally, suppose Merlin is honest and sends registers  $(X, Y)$  in the desired form, i.e.,  $X_j^i$  contains  $p_j(i) = \langle \Pi_j(i), \rho_j \rangle$  up to  $\alpha$  bits of precision, and  $Y_j^l$  contains  $\rho_j$ . Then, the expected value of the random variable  $n_j(i)$  is  $\mathbb{E}[n_j(i)] = k \langle \Pi_j(i), \rho_j \rangle$ , which is equal to  $k \cdot p_j(i)$  up to the error incurred by representing  $p_j(i)$  using  $\alpha$  bits of precision. In other words,

$$\left| \frac{\mathbb{E}[n_j(i)]}{k} - p_j(i) \right| < \frac{1}{2^\alpha} < \frac{1}{2p}. \quad (1)$$

We can hence upper bound the probability of rejecting in Step 3 by

$$\Pr \left[ \left| \frac{n_j(i)}{k} - p_j(i) \right| \geq \frac{1}{p} \right] < \Pr \left[ \left| \frac{n_j(i)}{k} - \frac{\mathbb{E}[n_j(i)]}{k} \right| \geq \frac{1}{2p} \right] \leq 2 \exp \left( -\frac{5p}{4} \right),$$

where the first inequality follows from Eq. (1) and the second from the Chernoff bound. Thus, Merlin passes Step 4 with probability exponentially close to 1.

We now turn to the final step. Since  $x \in A_{\text{yes}}$ , we know that the optimal distributions, denoted  $q_j := (\langle \Pi_j(1), \rho_j \rangle, \dots, \langle \Pi_j(r), \rho_j \rangle)$  for  $j \in [m]$ , obtained in Stage 1 of the original BellQMA protocol are now accepted in Stage 2 with probability at least  $2/3$ . However, in our case, Merlin was only able to specify each  $q_j$  up to  $\alpha$  bits of precision per entry as the distributions  $p_j$ . To analyze how this affects the probability of acceptance, let  $P_j$  and  $Q_j$  be diagonal operators with entries  $P_j(i, i) = p_j(i)$  and  $Q_j(i, i) = \langle \Pi_j(i), \rho_j \rangle$ , respectively. Letting  $\Lambda_{\text{accept}}$  denote the POVM element corresponding to outcome *accept* in Stage 2 of the BellQMA protocol, we thus bound the change in acceptance probability by:

$$\begin{aligned} \left| \text{Tr} \left[ \Lambda_{\text{accept}} \left( \bigotimes_{j=1}^m P_j - \bigotimes_{j=1}^m Q_j \right) \right] \right| &\leq \left\| \bigotimes_{j=1}^m P_j - \bigotimes_{j=1}^m Q_j \right\|_{\text{tr}} \\ &\leq \sum_{j=1}^m \|P_j - Q_j\|_{\text{tr}} \\ &= \sum_{j=1}^m \sum_{i=1}^r |p_j(i) - \langle \Pi_j(i), \rho_j \rangle| \\ &\leq \frac{mr}{2^{20n\alpha}}, \end{aligned}$$

where the first inequality follows from the fact that  $|\text{Tr}(AB)| \leq \|A\|_\infty \cdot \|B\|_{\text{tr}}$  and the second inequality follows from Lemma 2.1. Therefore, the probability of success for each of the  $q$  runs of the BellQMA protocol in Step 5 is at least

$$\left( \frac{2}{3} - \frac{mr}{2^{20n\alpha}} \right) > 0.6.$$

Since each run is independent, applying the Chernoff bound yields that Arthur accepts Merlin's proof in Step 5 with probability at least  $1 - 2 \exp(-0.02q)$ , as desired. There may be some error incurred in sampling, which can be assumed to be exponentially small so that the success probability of each run is still at least 0.6.



**Soundness.** We now prove that when  $x \in A_{\text{no}}$ , a dishonest Merlin can win with probability at most inverse polynomially bounded away from 1. To show this, we bound the probability of passing Step 4 by relating the quantity  $p_j(i)$  to the expected value of  $n_j(i)/k$ , and then apply the Markov bound. The desired relationship follows by observing first that the expected value of  $n_j(i)/k$  is precisely the probability of obtaining outcome  $i$  when measuring proof  $j$  of some (honest) unentangled strategy, followed by arguing that the distribution  $p_j$  must hence be far from this latter (honest) distribution if Merlin is to pass Step 5 with probability at least  $1/2$  (since  $x \in A_{\text{no}}$ ). Combining these facts, we find that Arthur detects a cheating Merlin with inverse polynomial probability in Step 4.

More formally, let the quantum register  $Y_j$  contain an arbitrary quantum state  $\sigma_j$  whose reduced states in registers  $Y_j^l$  for  $l \in [k]$  are given by  $\sigma_j(l)$ , and define

$$\xi_j := \frac{1}{k} \sum_{l=1}^k \sigma_j(l).$$

By the linearity of expectation, the expected value of the random variable  $n_j(i)/k$  is

$$\mathbb{E} \left[ \frac{n_j(i)}{k} \right] = \frac{1}{k} \sum_{l=1}^k \langle \Pi_j(i), \sigma_j(l) \rangle = \langle \Pi_j(i), \xi_j \rangle.$$

Our goal is to lower bound the expression

$$\Pr \left[ \left| \frac{n_j(i)}{k} - p_j(i) \right| \geq \frac{1}{p} \right]. \quad (2)$$

To achieve this, we first substitute  $p_j(i)$  above with a quantity involving  $\mathbb{E}[n_j(i)/k]$ , and then apply the Markov bound.

To relate  $\mathbb{E}[n_j(i)/k]$  to  $p_j(i)$ , we first remark that in order for Merlin to pass each run of Step 5 with probability exponentially close to 1, he must send probability distributions  $p_j$ , which are accepted by Stage 2 of the BellQMA verification with probability at least  $1/2$ . Let

$$q_j(i) := \langle \Pi_j(i), \xi_j \rangle.$$

Let us imagine a BellQMA protocol where the  $j$ -th Merlin sends  $\xi_j$  as his quantum proof. Since  $x \in A_{\text{no}}$ , by the soundness property of the BellQMA( $m$ ) proof system, the success probability of the Merlins is at most  $1/3$ . In other words, sampling outcomes from the probability distributions  $(q_j(1), \dots, q_j(r))$  and then running the second stage of the BellQMA verification will yield outcome *accept* with probability at most  $1/3$ . Also, observe that

$$\mathbb{E} \left[ \frac{n_j(i)}{k} \right] = q_j(i).$$

It follows that by letting  $P_j$  and  $Q_j$  be diagonal operators with the probability vectors  $p_j$  and  $q_j$  on their diagonals, respectively, and  $\Lambda_{\text{accept}}$  the POVM element corresponding to outcome *accept* in Stage 2 of the BellQMA protocol, we have

$$\frac{1}{10} < \left| \text{Tr} \left[ \Lambda_{\text{accept}} \left( \bigotimes_{j=1}^m P_j - \bigotimes_{j=1}^m Q_j \right) \right] \right| \leq \left\| \bigotimes_{j=1}^m P_j - \bigotimes_{j=1}^m Q_j \right\|_{\text{tr}} \leq \sum_{j=1}^m \|P_j - Q_j\|_{\text{tr}}.$$

Here, the (loose) lower bound of  $1/10$  comes from the following two observations. First, the distributions represented by the diagonal operators  $Q_j$ 's are derived from a BellQMA protocol and therefore achieve a success probability at most  $1/3$  by the soundness property of the BellQMA verification. Second, the distributions represented by the diagonal operators  $P_j$ 's have to achieve a success probability strictly greater than  $1/2$  per run to guarantee that Merlin wins Step 5 with probability exponentially close to 1. Combining these two, we get that the difference between the success probabilities obtained by distributions described by operators  $\{P_j : j \in [m]\}$  and  $\{Q_j : j \in [m]\}$  should be at least  $1/6$  modulo the error incurred due to finite precision when encoding the distributions  $p_j$ . The use of the constant  $1/10$  overcompensates for this precision error. Hence, there exists a  $j$  such that

$$\|P_j - Q_j\|_{\text{tr}} = \sum_{i=1}^r |p_j(i) - q_j(i)| \geq \frac{1}{10mr},$$

implying the existence of an  $i$  such that

$$|p_j(i) - q_j(i)| \geq \frac{1}{10mr}. \quad (3)$$

This is our desired relationship between  $p_j(i)$  and  $\mathbb{E}[n_j(i)/k] = q_j(i)$ . Note that the probability of picking pair  $(i, j)$  in Step 4 is  $1/mr$ .

We now substitute this relationship into Eq. (2) and apply the Markov bound. Specifically, choose  $i$  and  $j$  as in Eq. (3), and assume that  $p_j(i) > \langle \Pi_j(i), \xi_j \rangle$ . Then, we have

$$\Pr \left[ \left| \frac{n_j(i)}{k} - p_j(i) \right| < \frac{1}{p} \right] < \Pr \left[ \frac{n_j(i)}{k} - \mathbb{E} \left[ \frac{n_j(i)}{k} \right] > \frac{1}{10mr} - \frac{1}{p} \right] \leq 1 - \frac{1}{2p}.$$

The case of  $p_j(i) < \langle \Pi_j(1), \xi_j \rangle$  is similar. We conclude that a dishonest Merlin is caught in Step 4 with probability at least  $1/2p$ . Therefore, the probability that Arthur proceeds to Step 5 is upper bounded by

$$\left( \frac{1}{mr} \right) \left( 1 - \frac{1}{20mr} \right) + \left( 1 - \frac{1}{mr} \right) (1) = 1 - \frac{1}{20m^2r^2},$$

where the first term represents the case where Arthur selects the correct pair  $(i, j)$  to check, and the second term the complementary case, in which we assume the cheating prover can win with probability 1. Hence the overall success probability of a dishonest Merlin is at most  $1 - 1/20m^2r^2$ , which is bounded away from 1 by an inverse polynomial.

Finally, as mentioned before, since  $m$  and  $r$  are polynomially bounded functions, we have that the completeness is exponentially close to 1, while the soundness is bounded away from 1 by an inverse polynomial. By known amplification techniques for QMA protocols [KW00, MW05], one can amplify the completeness and soundness errors to be exponentially close to 0. This proves our desired containment.

## 5 Perfect parallel repetition for SepQMA(poly)

We now show Theorem 1.4, i.e., that the class SepQMA( $m$ ) admits perfect parallel repetition. Before we proceed, recall that the closed convex cone Sep  $(\mathcal{X}_1, \dots, \mathcal{X}_m)$  is defined to contain operators of the form

$$\sum_{i=1}^k P_1(i) \otimes \dots \otimes P_m(i),$$

where  $P_j(i) \in \text{Pos}(\mathcal{X}_j)$ , for every  $j \in [m]$  and  $i \in [k]$ . This is the cone of interest and it is known to be closed and convex with non-empty interior. Given  $C$  to be the measurement operator corresponding to outcome *accept*, the maximum success probability of the Merlins in any  $\text{QMA}(m)$  protocol can be written as the maximum of  $\langle \rho, C \rangle$ , where  $\rho$  is a density operator in  $\text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m)$ . By standard convexity argument, one can always assume that the maximum is achieved by a pure product state.

For the remainder of the section, it will be convenient for us to distinguish two instances of  $\text{SepQMA}(m)$  protocols as the *first* and *second* protocol. For the first  $\text{SepQMA}(m)$  protocol we can write the maximum acceptance probability as the optimal value of the primal problem in the following primal-dual pair (where the operator  $C_1$  is Arthur's POVM element corresponding to outcome *accept*):

<u>Primal problem (P<sub>1</sub>)</u>	<u>Dual problem (D<sub>1</sub>)</u>
maximize: $\langle \rho_1, C_1 \rangle$	minimize: $t_1$
subject to: $\text{Tr}(\rho_1) = 1,$	subject to: $t_1 \mathbb{1}_{\mathcal{X}} = C_1 + W_1,$
$\rho_1 \in \text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m),$	$W_1 \in \text{Sep}^*(\mathcal{X}_1, \dots, \mathcal{X}_m),$

where  $\mathcal{X}$  denotes  $\mathcal{X}_1 \otimes \dots \otimes \mathcal{X}_m$ . The use of “maximum” and “minimum” is justified in the above programs since

$$\bar{\rho}_1 = \frac{\mathbb{1}_{\mathcal{X}}}{\dim(\mathcal{X})} \quad \text{and} \quad (\bar{t}_1, \bar{W}_1) = (2, 2\mathbb{1}_{\mathcal{X}} - C_1)$$

are strictly feasible solutions for (P<sub>1</sub>) and (D<sub>1</sub>), respectively [GB02, GB03, GB05]. Hence, by Theorem 2.6, strong duality holds for both problems, i.e., both problems attain an optimal solution and the optimal values are the same. We can similarly formulate the acceptance probability of the second protocol as

<u>Primal problem (P<sub>2</sub>)</u>	<u>Dual problem (D<sub>2</sub>)</u>
maximize: $\langle \rho_2, C_2 \rangle$	minimize: $t_2$
subject to: $\text{Tr}(\rho_2) = 1,$	subject to: $t_2 \mathbb{1}_{\mathcal{Y}} = C_2 + W_2,$
$\rho_2 \in \text{Sep}(\mathcal{Y}_1, \dots, \mathcal{Y}_m),$	$W_2 \in \text{Sep}^*(\mathcal{Y}_1, \dots, \mathcal{Y}_m),$

where  $\mathcal{Y}$  denotes  $\mathcal{Y}_1 \otimes \dots \otimes \mathcal{Y}_m$ . Since we are considering  $\text{SepQMA}$  protocols it holds that

$$C_1 \in \text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m) \quad \text{and} \quad C_2 \in \text{Sep}(\mathcal{Y}_1, \dots, \mathcal{Y}_m).$$

Given the two cone programs above, the maximum acceptance probability of the two-fold repetition of the protocol can hence be expressed as

<u>Primal problem (P)</u>	<u>Dual problem (D)</u>
maximize: $\langle \rho, C_1 \otimes C_2 \rangle$	minimize: $t$
subject to: $\text{Tr}(\rho) = 1,$	subject to: $t \mathbb{1}_{\mathcal{X} \otimes \mathcal{Y}} = C_1 \otimes C_2 + W,$
$\rho \in \text{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$	$W \in \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m).$

Note that the operators  $\rho$  and  $W$  are elements of  $\text{Herm}(\mathcal{X}_1 \otimes \dots \otimes \mathcal{X}_m \otimes \mathcal{Y}_1 \otimes \dots \otimes \mathcal{Y}_m)$ .

To show Theorem 1.4, observe that if  $\rho_1$  and  $\rho_2$  are any respective optimal solutions of (P<sub>1</sub>) and (P<sub>2</sub>), then  $\rho_1 \otimes \rho_2$  is a feasible solution of (P). Therefore the optimal value of (P) is at least the product of the optimal values of (P<sub>1</sub>) and (P<sub>2</sub>). It remains to show that in fact *no* other strategy for

the prover can perform better than this honest strategy. To do so, we demonstrate a dual feasible solution for (D) attaining this same objective value.

More formally, let  $(t_1, W_1)$  and  $(t_2, W_2)$  be respective dual optimal solutions of  $(D_1)$  and  $(D_2)$ . By strong duality,  $t_1$  is the optimal value of  $(P_1)$  and  $t_2$  is the optimal value of  $(P_2)$ . We show that  $t_1 \cdot t_2$  is an upper bound on the optimal value of  $(P)$  by exhibiting a solution  $(t_1 \cdot t_2, W)$  which is feasible in  $(D)$ , for some  $W \in \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$ . We first prove the following useful lemma.

**Lemma 5.1.** *For complex Euclidean spaces  $\mathcal{X}_1, \dots, \mathcal{X}_m$  and  $\mathcal{Y}_1, \dots, \mathcal{Y}_m$ , the following two containments hold:*

- $\text{Sep}^*(\mathcal{X}_1, \dots, \mathcal{X}_m) \otimes \text{Sep}(\mathcal{Y}_1, \dots, \mathcal{Y}_m) \subseteq \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$ , and
- $\text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m) \otimes \text{Sep}^*(\mathcal{Y}_1, \dots, \mathcal{Y}_m) \subseteq \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$ .

*Proof.* We prove the first condition as the second is nearly identical. Fix  $W \in \text{Sep}^*(\mathcal{X}_1, \dots, \mathcal{X}_m)$  and  $C \in \text{Sep}(\mathcal{Y}_1, \dots, \mathcal{Y}_m)$ . Then for  $S \in \text{Sep}(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$ , we have

$$\langle W \otimes C, S \rangle = \langle W, \text{Tr}_{\mathcal{Y}} [S(\mathbb{1}_{\mathcal{X}} \otimes C)] \rangle \geq 0,$$

if  $\text{Tr}_{\mathcal{Y}} [S(\mathbb{1}_{\mathcal{X}} \otimes C)] \in \text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m)$ . Therefore, it suffices to prove that  $\text{Tr}_{\mathcal{Y}} [S(\mathbb{1}_{\mathcal{X}} \otimes C)] \in \text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m)$ . To this end, let

$$S = \sum_{i=1}^k \bigotimes_{l=1}^m \rho_i(l) \quad \text{and} \quad C = \sum_{j=1}^{k'} \bigotimes_{l=1}^m \sigma_j(l),$$

where  $\rho_i(l) \in \text{Pos}(\mathcal{X}_l \otimes \mathcal{Y}_l)$  and  $\sigma_j(l) \in \text{Pos}(\mathcal{Y}_l)$  for all  $i \in [k]$ ,  $j \in [k']$ , and  $l \in [m]$ . Now we can write  $\text{Tr}_{\mathcal{Y}} [S(\mathbb{1}_{\mathcal{X}} \otimes C)]$  as

$$\text{Tr}_{\mathcal{Y}} \left[ \left( \sum_{i=1}^k \bigotimes_{l=1}^m \rho_i(l) \right) \left( \mathbb{1}_{\mathcal{X}} \otimes \sum_{j=1}^{k'} \bigotimes_{l=1}^m \sigma_j(l) \right) \right] = \sum_{i=1}^k \sum_{j=1}^{k'} \bigotimes_{k=1}^m \text{Tr}_{\mathcal{Y}_k} [\rho_i(k) (\mathbb{1}_{\mathcal{X}_k} \otimes \sigma_j(k))],$$

which is clearly in  $\text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m)$ . This concludes the proof.  $\square$

We now use Lemma 5.1 to construct two operators in  $\text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$ , the appropriate convex combination of which is the dual feasible solution we are seeking. Specifically, observe first that since for the two instances of the  $\text{SepQMA}(m)$  protocol, we have  $C_1 \in \text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m)$  and  $C_2 \in \text{Sep}(\mathcal{Y}_1, \dots, \mathcal{Y}_m)$ , and since  $\mathbb{1}_{\mathcal{X}}$  and  $\mathbb{1}_{\mathcal{Y}}$  are fully separable operators, it follows that

$$t_1 \mathbb{1}_{\mathcal{X}} + C_1 \in \text{Sep}(\mathcal{X}_1, \dots, \mathcal{X}_m) \quad \text{and} \quad t_2 \mathbb{1}_{\mathcal{Y}} + C_2 \in \text{Sep}(\mathcal{Y}_1, \dots, \mathcal{Y}_m)$$

for all  $t_1, t_2 \geq 0$ . Using Lemma 5.1, we thus obtain operators

$$(t_1 \mathbb{1}_{\mathcal{X}} - C_1) \otimes (t_2 \mathbb{1}_{\mathcal{Y}} + C_2) \in \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m) \quad (4)$$

and

$$(t_1 \mathbb{1}_{\mathcal{X}} + C_1) \otimes (t_2 \mathbb{1}_{\mathcal{Y}} - C_2) \in \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m), \quad (5)$$

where  $t_1 \mathbb{1}_{\mathcal{X}} - C_1 \in \text{Sep}^*(\mathcal{X}_1, \dots, \mathcal{X}_m)$  by the constraints of  $(D_1)$ , and similarly for  $t_2 \mathbb{1}_{\mathcal{Y}} - C_2$ . Since  $\text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m)$  is a convex cone, it follows that the average of Eqs. (4) and (5) yields the desired operator

$$W := t_1 \cdot t_2 \mathbb{1}_{\mathcal{X} \otimes \mathcal{Y}} - C_1 \otimes C_2 \in \text{Sep}^*(\mathcal{X}_1 \otimes \mathcal{Y}_1, \dots, \mathcal{X}_m \otimes \mathcal{Y}_m).$$

We conclude that  $(t_1 \cdot t_2, W)$  is a feasible solution of the dual problem (D) with objective value  $t_1 \cdot t_2$  as desired. This concludes the proof of Theorem 1.4.

We note that there are instances of QMA(poly) protocols, which are not SepQMA(poly) protocols, that admit perfect parallel repetition. Although this fact is known in the literature (see Harrow and Montanaro [HM10] for details), we provide a concrete example below.

First, note that the maximum acceptance probability of Arthur in a QMA( $m$ ) protocol is upper bounded by  $\|C\|_\infty$ , where  $C$  is the accepting measurement operator. Now, consider the two-qubit POVM operator

$$C := \frac{1}{2} |00\rangle \langle 00| + \frac{1}{2} |\Psi^+\rangle \langle \Psi^+|,$$

where

$$|\Psi^+\rangle := \frac{1}{\sqrt{2}} |01\rangle + \frac{1}{\sqrt{2}} |10\rangle.$$

We can easily check that  $C$  has two eigenvalues, 0 and  $1/2$ , and two principal eigenvectors  $|00\rangle$  and  $|\Psi^+\rangle$ , one of which is a product state. It follows that the maximum acceptance probability is  $1/2$ . By the multiplicative property of the infinity-norm under tensor products, it holds that the maximum acceptance probability of the  $k$ -fold repetition is exactly  $1/2^k$ .

We now argue that  $C$  is not a separable operator. Suppose for the sake of contradiction that  $C$  can be written as

$$\sum_{i=1}^n \rho_i \otimes \sigma_i$$

for some  $\rho_i, \sigma_i \in \text{Pos}(\mathbb{C}^2)$ . Then we have

$$0 = \langle C, |11\rangle \langle 11| \rangle = \sum_{i=1}^n \langle 1 | \rho_i | 1 \rangle \langle 1 | \sigma_i | 1 \rangle,$$

which implies  $\rho_i |1\rangle = 0$  or  $\sigma_i |1\rangle = 0$  for all  $i \in [n]$ . This leads to the contradiction

$$\frac{1}{4} = \langle C, |01\rangle \langle 10| \rangle = \sum_{i=1}^n \langle 1 | \rho_i | 0 \rangle \langle 0 | \sigma_i | 1 \rangle = 0.$$

Alternatively, one can show that  $C$  is not separable by observing that  $C$  has a non-positive partial transpose [Per96, HHH96].

## 6 Conclusions and open problems

In this paper, we have studied three variants of multi-prover quantum Merlin-Arthur proof systems. We first showed that a system with polynomially many provers is indeed strictly more powerful than a single prover system if messages are restricted to be logarithmic in length, unless  $\text{BQP} = \text{MQA}$ . We next showed that polynomially many provers do not provide additional expressive power over a single prover in the setting where the verifier is restricted to first applying unentangled and non-adaptive measurements with at most a polynomial number of outcomes per proof. Both of these questions make steps towards understanding the major open question of whether QMA with polynomially many provers is more powerful than QMA. Finally, we used cone programming duality to give an alternate proof of the fact that perfect parallel repetition holds whenever a QMA verifier's POVM element corresponding to *accept* is a fully separable operator.



A consequence of our first result is that the two variants of the class  $\text{QMA}(\text{poly})$ , where Merlins send logarithmic-size proofs and Merlins send constant-size proofs are equal. A natural question concerning our first result is to understand the expressive power of the variant of  $\text{QMA}(\text{poly})$ , where Merlins are restricted to send  $\text{poly} \log(|x|)$  qubits to Arthur. Another open question concerning the results presented in this paper is the relationship between  $\text{BellQMA}(\text{poly})$  and  $\text{QMA}$ . We believe that understanding the complexity of  $\text{BellQMA}$  protocols, or more generally  $\text{LOCC-QMA}$  protocols, will shed new light on the bigger question pertaining to  $\text{QMA}(2)$  and  $\text{QMA}$ . Another avenue of interest is to find further applications of the cone programming characterization of multi-prover quantum Merlin-Arthur proof systems. A straightforward question concerning the parallel repetition result presented in this paper is to investigate whether cone programming duality can be used to analyze the product state test in the Ref. [HM10]. Another question one can ask is to find other classes of  $\text{QMA}(m)$  protocols that admit a perfect parallel repetition theorem.

## Acknowledgements

We thank Richard Cleve, Tsuyoshi Ito, Iordanis Kerenidis, Ashwin Nayak, Oded Regev, and Lev-ent Tunçel for insightful discussions. We also thank the EU-Canada Exchange Program and LI-AFA, Paris for their hospitality, where part of this work was completed. SG acknowledges support from NSERC, NSERC MSFSS, the David R. Cheriton Graduate Scholarship program, the President's Graduate Scholarship, and CIFAR. JS acknowledges support from NSERC, MITACS, ERA (Ontario), and the President's Graduate Scholarship. SU acknowledges support in parts from CIFAR, MITACS, NSERC, Ontario's Ministry of Research and Innovation, QuantumWorks, the U.S. A.R.O., David R. Cheriton Graduate Scholarship, and the Mike and Ophelia Graduate Fellowship.

## References

- [Aar06] S. Aaronson.  $\text{QMA}/\text{qpoly}$  is contained in  $\text{PSPACE}/\text{poly}$ : De-Merlinizing quantum protocols. In *Proceedings of the 21st Annual IEEE Conference on Computational Complexity*, pages 273–286, 2006.
- [AB09] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.
- [ABD<sup>+</sup>09] S. Aaronson, S. Beigi, A. Drucker, B. Fefferman, and P. Shor. The power of unentanglement. *Theory of Computing*, 5:1–42, 2009.
- [ABOBS08] D. Aharonov, M. Ben-Or, F. Brandão, and O. Sattath. The pursuit for uniqueness: Extending Valiant-Vazirani theorem to the probabilistic and quantum settings. Available at arXiv.org e-Print quant-ph/0810.4840v1, 2008.
- [AGIK09] D. Aharonov, D. Gottesman, S. Irani, and J. Kempe. The power of quantum systems on a line. *Communications in Mathematical Physics*, 287(1):41–65, 2009.
- [AK07] S. Aaronson and G. Kuperberg. Quantum versus classical proofs and advice. *Theory of Computing*, 3:129–157, 2007.
- [ALM<sup>+</sup>98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

- [AN02] D. Aharonov and T. Naveh. Quantum NP - a survey. Available at arXiv.org e-Print quant-ph/0210077v1, 2002.
- [AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.
- [BCY11] F. Brandão, M. Christandl, and J. Yard. A quasipolynomial-time algorithm for the quantum separability problem. Available at arXiv.org e-Print quant-ph/1011.2751v2, 2011.
- [Bei08] S. Beigi. NP vs  $\text{QMA}_{\log}(2)$ . Available at arXiv.org e-Print quant-ph/0810.5109v1, 2008.
- [Bra06] S. Bravyi. Efficient algorithm for a quantum analogue of 2-SAT. Available at arXiv.org e-Print quant-ph/0602108, 2006.
- [Bra08] F. Brandão. *Entanglement Theory and the Quantum Simulation of Many-Body Physics*. PhD thesis, Imperial College London, London, 2008. Available at arXiv.org e-Print quant-ph/1011.2751v2.
- [BS07] S. Beigi and P. Shor. On the complexity of computing zero-error and Holevo capacity of quantum channels. Available at arXiv.org e-Print quant-ph/0709.2090, 2007.
- [BT09] H. Blier and A. Tapp. All languages in NP have very short quantum proofs. In *Proceedings of the 3rd International Conference on Quantum, Nano and Micro Technologies*, pages 34–37, 2009. Available at arXiv.org e-Print quant-ph/0709.0738v2, first posted in 2007.
- [BV97] E. Bernstein and U. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.
- [CD10] J. Chen and A. Drucker. Short multi-prover quantum proofs for SAT without entangled measurements. Available at arXiv.org e-Print quant-ph/1011.0716v2, 2010.
- [Coo71] S. Cook. The complexity of theorem proving procedures. In *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, pages 151–158, 1971.
- [CSUU08] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. *Computational Complexity*, 17(2):282–299, 2008.
- [GB02] L. Gurvits and H. Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6):062311, 2002.
- [GB03] L. Gurvits and H. Barnum. Separable balls around the maximally mixed multipartite quantum states. *Physical Review A*, 68(4):042312, 2003.
- [GB05] L. Gurvits and H. Barnum. Better bound on the exponent of the radius of the multipartite separable ball. *Physical Review A*, 72(3):032322, 2005.
- [Gha10] S. Gharibian. Strong NP-hardness of the quantum separability problem. *Quantum Information and Computation*, 10(3&4):343–360, 2010. Available at arXiv.org e-Print quant-ph/0810.4507, 2008.

- [Gur03] L. Gurvits. Classical deterministic complexity of Edmonds problem and quantum entanglement. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 10–19, 2003.
- [Gut09] G. Gutoski. *Quantum strategies and local operations*. PhD Thesis, University of Waterloo, 2009. Available at arXiv.org e-Print quant-ph/1003.0038.
- [HHH96] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996.
- [HHHH09] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Review in Modern Physics*, 81(2):865–942, 2009.
- [HM10] A. Harrow and A. Montanaro. An efficient test for product states, with applications to quantum Merlin-Arthur games. In *Proceedings of the 51st IEEE Annual Symposium on Foundations of Computer Science*, pages 633–642, 2010.
- [Ioa07] L. Ioannou. Computational complexity of the quantum separability problem. *Quantum Information and Computation*, 7(4):335–370, 2007.
- [JGL10] S. Jordan, D. Gosset, and P. Love. QMA-complete problems for stoquastic Hamiltonians and Markov matrices. *Physical Review A*, 81(3):032331, 2010.
- [JW06] D. Janzing and P. Wocjan. BQP-complete problems concerning mixing properties of classical random walks on sparse graphs. Available at arXiv.org e-Print quant-ph/0610235v2, 2006.
- [KKR06] J. Kempe, A. Kitaev, and O. Regev. The complexity of the local Hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006.
- [KMY03] H. Kobayashi, K. Matsumoto, and T. Yamakami. Quantum Merlin Arthur proof systems: Are multiple Merlins more helpful to Arthur? In *Proceedings of the 14th International Symposium on Algorithms and Computation*, pages 189–198, 2003. Volume 2906 of *Lecture Notes in Computer Science*, Springer.
- [KR03] J. Kempe and O. Regev. 3-local Hamiltonian is QMA-complete. *Quantum Information and Computation*, 3(3):258–264, 2003.
- [KRT10] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010.
- [KSV02] A. Kitaev, A. Shen, and M. Vyalys. *Classical and Quantum Computation*. American Mathematical Society, 2002.
- [KW00] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 608–617, 2000.
- [LCV07] Y.-K. Liu, M. Christandl, and F. Verstraete. Quantum computational complexity of the N-representability problem: QMA complete. *Physical Review Letters*, 98(11):110503, 2007.
- [Lev73] L. Levin. Universal search problems. *Problems of Information Transmission*, 9(3):265–266, 1973. In Russian.

- [Liu06] Y.-K. Liu. Consistency of local density matrices is QMA-complete. In *Proceedings of the 10th International Workshop on Randomization and Computation*, pages 438–449, 2006. Volume 4110 of *Lecture Notes in Computer Science*, Springer.
- [MW05] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005.
- [NC00] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [OT08] R. Oliveira and B. Terhal. The complexity of quantum spin systems on a two-dimensional square lattice. *Quantum Information and Computation*, 8(10):0900–0924, 2008.
- [Per96] A. Peres. Separability criterion for density matrices. *Physical Review Letters*, 77:1413, 1996.
- [Ros11] B. Rosgen. Testing non-isometry is QMA-complete. In *Proceedings of the 5th Conference on Theory of Quantum Computation, Communication, and Cryptography*, pages 63–76, 2011. Volume 6519 of *Lecture Notes in Computer Science*, Springer.
- [SV09] N. Schuch and F. Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5:732–735, 2009.
- [TW08] L. Tunçel and H. Wolkowicz. Strong duality and minimal representations for cone optimization. Technical Report CORR 2008-07, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, August 2008 (revised: December 2008).
- [Wat02] J. Watrous. Limits on the power of quantum statistical zero-knowledge. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 459–468, 2002.
- [Wat09] J. Watrous. *Encyclopedia of Complexity and System Science*, chapter Quantum Computational Complexity. Springer, 2009.
- [WMN10] T.-C. Wei, M. Mosca, and A. Nayak. Interacting boson problems can be QMA-hard. *Physical Review Letters*, 104(4):040501, 2010.
- [WY08] P. Wocjan and J. Yard. The Jones polynomial: Quantum algorithms and applications in quantum complexity theory. *Quantum Information and Computation*, 8(1&2):0147–0180, 2008.